Exchange of Bibliographic Information in Machine-readable Form

LUCIA J. RATHER

THE TRANSFER OF bibliographic information between agencies is not a new idea. It began with the earliest printed catalogs and became a well-established operation in the early part of the twentieth century with the sale of Library of Congress catalog cards. The need for standards to aid in the efficient transfer of bibliographic information became apparent early; the first such major standard may have been the specification for the 3 × 5 inch catalog card.

The development of networks for the transfer of bibliographic information in machine-readable form from one agency to another in the 1970s is a change in form rather than in substance. The necessity to use complex electronic equipment and the desire to utilize bibliographic records without extensive manual modification have made standardization even more significant. In 1970, Wigington and Wood stated that if "a national program for information transfer has as its objectives the development of a coherent system for the efficient, effective, and economic transfer of information, then the need for a standardization program which is much more extensive than anything available to date becomes obvious."[1]

Standards are created today at many different levels. Schmierer has noted five groups of institutions involved in the creation of standards.[2] At the international level is the International Organization for Standardization (ISO), founded in 1946. ISO works with national and regional standards organizations to develop recommendations and standards in all areas of technology. At the regional level are organizations such as the Pan American Standards Coordinating Committee (COPANT) and the European Committee for Standardization (CEN),

Lucia J. Rather is Assistant Director (Cataloging), Processing Department, Library of Congress.

which generally work with ISO rather than developing standards on their own. Most of the larger countries contain national standards organizations. The U.S. agency is the American National Standards Institute (ANSI). ANSI works with both private and governmental groups in the establishment of national standards. Standards are also developed at the associational level by groups recognized as authorities for industry-wide interest. One of the best known of these groups is the American Society for Testing and Materials. At the fifth level are government agencies, companies, and individuals; the prime example in the United States is the National Bureau of Standards.

The main groups in the United States concerned with library standards are ANSI Subcommittee Z39 on Library Work and Documentation, the Library of Congress, and the Committee on Technical Standards for Library Automation (TESLA) of the Information Science and Automation Division of the American Library Association. TESLA, like the Library of Congress (LC), is not in itself a standards-making organization. Instead, it promotes "participation in the standards process at the membership level" so that standards developed by organizations such as ANSI will reflect library needs.[3]

COMMUNICATION OF BIBLIOGRAPHIC DATA IN MACHINE-READABLE FORM

Manual methods of communication of bibliographic information are enhanced by the use of standards. When automated methods are used, standards become critical. This may be explained by the fact that one of the prime economic justifications of automated systems lies in the fact that they allow the manipulation of data without expensive manual intervention. Computer programs are very intolerant of data variations. People, on the other hand, have the superior ability of making allowances for data variations when they process records. People can also make bibliographic decisions based on implicit information. They do not have to be told explicitly that a given record is in French, that Paris is a city in France, or that "Bibliothèque nationale" is a corporate name. For rapid, accurate processing, the computer must be told this explicitly, and if records from more than one source are to be processed, the information must be conveyed in exactly the same way.

The requirements for standardization in order to transmit bibliographic records in machine-readable form fall into seven areas: (1) bibliographic description; (2) standardization of headings, subject

terms, and other access points; (3) transliteration; (4) character sets; (5) formats; (6) codes; and (7) item identification. Some of these areas have been described more fully elsewhere in this issue and will be touched on only briefly here.

BIBLIOGRAPHIC DESCRIPTION

An examination of the national bibliographies issued prior to 1972 indicates wide variation in the rules for bibliographic description. Order of information varied, as did punctuation between data elements. Published in 1974, the *International Standard Bibliographic Description for Monographic Publications* (ISBD(M))[4] was the first attempt to standardize description on an international scale. ISBD(M) has been widely adopted by most of the European countries and formed the basis for a revision of Chapter 6 of the *Anglo-American Cataloging Rules* (AACR).[5]

At the Library of Congress, which has had a policy of accepting the bibliographic description provided by national bibliographic agencies participating in its shared cataloging program, the adoption of ISBD(M) has had a substantial impact. LC uses a complex computer program called Format Recognition[6] to convert records to machine-readable form. Format Recognition processes the entire bibliographic record and assigns tags, indicators and subfield codes based on an analysis of the data (including key words and punctuation patterns in the record). The original program was written for materials in English using the AACR rules. The expansion of the program to handle foreign-language cataloging was made possible only by the adoption of the standardized punctuation patterns required by ISBD(M). This means that one version of the program can handle records in any language as long as the ISBD(M) rules are followed. The savings in the costs of conversion to machine-readable form are so substantial that LC now requires any material in the shared cataloging program going into MARC (Machine-readable Cataloging) to be modified to follow ISBD(M) punctuation patterns by the LC catalogers—if the foreign agency supplying the catalog record has not already done so. Some problems do exist in the current system, the most difficult one being that the different national agencies do not apply ISBD(M) in exactly the same fashion. This requires some modification of the catalog record by the LC catalogers, and in some cases causes problems in the Format Recognition programs.

STANDARDIZATION OF HEADINGS

There is far less standardization in the area of name headings than there is in bibliographic description. The Paris Principles provided some guidance, and various national cataloging codes such as AACR have been developed based on these principles. However, even if international agreement could be reached on rules for choice and form of heading, it remains likely that the headings created by different agencies would vary. This problem arises from two sources. First is the difficulty of devising rules which will be consistently interpreted alike by two different catalogers. This situation results both from the way the rules are written and from the infinite variety of names of people, societies, meetings, institutions, governments, etc. The second problem derives from the rules which specify what form of name is to be used but allow any additional information necessary to distinguish one name from another to be added. This kind of decision can be made only by comparing a given name against an authority file. Because there is no single authority file in existence, there is a wide divergence in the forms of headings established for a given name by catalogers at different agencies.

Libraries in the United States have long searched the National Union Catalog (NUC)[7] to determine the forms of names established at LC. In response to this need, LC has recently begun publishing a list of *Name Headings With References*[8] which includes all new names established during a given quarter. The economic problems of cumulation and the delay with which the list appears remain. Ease of cumulation and speed of publication seem to depend on the development of an automated authority system. The New York Public Library and the Washington State Library have been pioneers in this area, using machine-readable authority files for the production of book catalogs. The National Library of Canada also has its authority files in machine-readable form. The Library of Congress has developed a format for the handling of authority data and plans to begin input of new name headings in 1976. These headings will be available initially in machine-readable form on tape from the MARC Distribution Service. A committee organized by the Advisory Group on National Bibliographic Control is currently considering development of an authority format for names to satisfy the requirements of library publishers and abstracting and indexing communities. The committee is beginning its work with a review of the LC authority format.

The ideal solution would be to handle name information via an on-line system which could be searched or updated by multiple users,

with one agency serving as the final authority. A centralized system would allow all users to obtain up-to-date information on any given heading immediately. While such a system eventually may be possible, many problems remain to be solved, including quality control, duplication of headings, and varying requirements of participating agencies.

The question of standardization of subject headings and subject classification presents even more problems than do name systems. At this time, there is no one classification or subject heading system that can be considered truly international in scope. The Universal Decimal Classification (UDC) is actually an umbrella term for a group of national classification systems, all built on the same principles but varying substantially in terms of development and application. UDC is widely used in Europe but scarcely at all in the United States. The Library of Congress and Dewey Decimal classifications are the most common systems found in the United States. In the area of subject headings, there is even less agreement. In the United States and Canada, LC subject headings are widely used in public and research libraries. However, specialized systems, such as the Medical Subject Headings (MESH), are used by special libraries. There is currently no evidence that a translation can automatically be made from any of the specialized systems to the LC system or from one classification system to another. In the international area, this problem is compounded by language.

The Preserved Context Indexing System (PRECIS) developed by Derek Austin at the British National Bibliography is designed for use in an automated system, although it can also be used manually. PRECIS uses a series of subject terms to provide index entries to a record. The British Library is now planning to modify the system to handle translingual indexing. Under the envisioned scheme, thesauri will be developed in different languages, with each term given a number. The machine record will carry a set of numbers which can be translated into the appropriate subject terms in any given language. Whether the linguistic complexities of the various European languages will permit such equivalences, and whether this system will ever be widely used in the United States, is open to conjecture.

Until the subject approach can be standardized, agencies receiving machine-readable records from other agencies using different subject systems will be forced to perform costly subject analysis before the record can be used.

TRANSLITERATION

There are currently many transliteration and romanization systems in use. Some are letter-for-letter reversible; others rely on equivalences in pronunciation. The systems developed by the Library of Congress and the American Library Association (ALA) are widely used by libraries in the United States. Other systems include those developed by ANSI and by ISO. Completely reversible systems are frequently unattractive to scholars, but do offer the best opportunity in automated systems to allow conversion between the vernacular and transliterated forms. If complete reversibility were used, it might be possible to receive a record with Cyrillic characters and display it in Roman alphabet characters. Until there is widespread agreement on transliteration, however, communication of bibliographic records for non-Roman materials will be difficult.

CHARACTER SETS

Bibliographic records require the use of many different special characters and diacritical marks in addition to the normal alphabetic letters and punctuation marks. Each character must be determined and a method defined to input the character, store it, and display it. The early computer systems handled only uppercase letters and punctuation. Lowercase letters became available in the early 1950s and a standardized character set called ASCII was developed by ANSI in 1968.[9] The MARC Pilot Project used a character set with upper- and lowercase letters and a limited number of the more common diacritical marks. In 1968, a 175-character set was developed by the Library of Congress and the American Library Association handling 37 Roman alphabet languages and 34 additional languages in romanized form.[10] The LC/ALA set is an expanded form of ASCII and is used in the MARC Distribution Service. A number of hardware devices are now available with the capability of handling the 175 characters, including a computer print train, several cathode ray tube terminals, and a photocomposition system. The LC/ALA set was enlarged slightly by the Ohio College Library Center (OCLC) and is utilized in that system also.

As MARC systems were developed in other countries, the LC/ALA set was utilized, but it eventually became clear that modifications were needed if it were to satisfy foreign users. In 1972, ISO established a working group (ISO TC 46/SC 4/WG 1) to develop an international extended Roman character set and to investigate non-Roman char-

acter sets as well. Working Group 1 has now prepared draft proposals for an extended Latin set, a Greek set, and an extended Cyrillic set.[11] These proposals are currently under review by the various national standards organizations. Still under development are sets to handle mathematical characters and special characters used in the African languages.

In the United States, a system to handle Hebrew has been developed at the New York Public Library and work is in progress on a Hebrew set at OCLC. The Library of Congress is currently investigating the problem of carrying non-Roman alphabets in its automated system.

FORMATS

A format is the container which carries both data and data identifiers in a machine system. Data must be identified explicitly if it is to be processed and manipulated. One of the earliest efforts in this area was the report prepared by Lawrence Buckland in 1964 for the Council on Library Resources.[12] This was followed in 1965 by a planning memorandum issued by LC[13] detailing the feasibility of storing cataloging information in machine-readable form, which served as the basis for the first MARC format[14] used in the MARC Pilot Project. One of the purposes of the Pilot Project was to examine the format in use to determine needed modification.

The MARC II format[15] was developed as a result of an extensive examination of the MARC Pilot Project. Librarians from all over the United States met with staff members of the Library of Congress during a period of two years, and the format at various stages was reviewed by ALA committees. In 1968, it was adopted as a "standard" by ALA and serves today as the primary basis for the communication of cataloging information in machine-readable form.

The MARC format may be considered to be made up of two parts: the structure and the content designators. The structure is the framework of the format. It specifies certain control information needed for communication of information and defines the layout of the data fields. The structure is generalized and may be used for any type of bibliographic data. The structure of the format has now been adopted as a national standard by the American National Standards Institute as ANSI Z39.2 1971; it has also been adopted as an international standard by the International Organization for Standardization as ISO 2709.[16] Since the adoption of ISO 2709, several proposals have been made to modify it, but in such a way that the original structure

would still be valid for international use. No changes have yet received widespread support, but the format was scheduled for review in May 1976 at a meeting in Brussels of ISO TC 46/SC 4/WG 4.

The content designators are the tags, indicators, and subfield codes used to identify the data in the record. For different types of bibliographic materials (e.g., books, serials, maps, etc.), the data elements—and hence the content designators—may be different. The original MARC II format was designed to handle book material. MARC formats have been designed since then to cover serials, films, maps, manuscripts, and music.[17] The principle was established that if data elements were the same across forms of material, the same content designators would be used. A composite version of all the MARC formats is now in preparation.[18] The Library of Congress has also developed an authority format[19] using the MARC structure and a simplified format[20] to handle in-process materials.

The first uses of the MARC format were in the distribution of catalog records by the Library of Congress. The MARC Distribution Services have primarily been confined to current cataloging. The books service began in 1969 with English-language material. In 1973 it was expanded to French; in 1975 to German, Spanish and Portuguese; and in 1976 to Italian, Dutch and Scandinavian. In 1977, it will include all Roman alphabet cataloging. Materials in non-Roman alphabets will be added in the next three years so that by 1979, all LC current cataloging should be in machine-readable form. Techniques for handling non-Roman alphabets are still under consideration. The books data base now includes more than 630,000 records. The films distribution service, which began in 1972, now includes more than 30,000 records. The serials and maps services began in 1973 and include 25,000 and 21,000 titles, respectively. Records in these last three services cover all languages with non-Roman data carried in romanized form.

The format has subsequently been used in more or less standard form by other libraries and bibliographic agencies in the United States. This has naturally promoted the idea that these MARC records created outside LC should become part of a centralized data base to the greater good of the library community. The problems appeared to be twofold. In some cases, the cataloging was thought to be less than reliable; in other cases, the full array of content designators was not used. In 1975, the Library of Congress, with the support of the Council on Library Resources, began a new project called Cooperative MARC (COMARC) designed to test the feasibility of

building such a centralized cooperative data base. Libraries using the full MARC format were invited to contribute records based on LC cataloging (taken from an LC printed card) to the Library of Congress. LC agreed to take these records, eliminate duplicates, compare them against the official catalog card in the library's official catalog, and update any of the access points (headings, subjects, call number, title) where these had been changed. LC would then reissue them as part of its MARC Distribution Service with the proviso that libraries contributing to COMARC would get all the records free of charge.

When the first libraries applied to LC, it became apparent immediately that very few libraries were using the MARC format in its entirety. A compromise position was finally taken under which some of the coded fixed-field information was not required, with this information signaled by a "fill character." In addition, some variable fields (such as National Bibliography Number) were made optional, but the principle was established that when a field was included, it should be defined by the full set of content designators. A second problem arose from the fact that some libraries were modifying the record by changing the edition statement or the imprint but retaining the LC card number. These records will be deleted from the system. Because the economics of the project were based on the premise that LC would check out only the access points, this raised the possibility that some very nonstandard cataloging might remain in collation, notes fields, etc. Whether this will prove to be a serious problem to other libraries using these records remains to be determined.

As of April 1976, three agencies were actively participating in the project: the Washington State Library, the University of Chicago, and the Information Dynamics Corporation. Three others, Cornell University, Boston Theological Institute, and Northwestern University had been accepted by that date and were expected to begin submitting tapes shortly. The first tapes, containing approximately 6,000 updated COMARC records, became available through the MARC Distribution Service in May 1976. The proof of the success or failure of this project will come through the ability of outside libraries to use the records. This will not be known until the tapes have been in distribution for some time, but if it is successful, LC hopes to continue the project on a wider scale, perhaps including non-LC cataloging as well.

The primary criticism of the MARC formats in the United States has been that they are too complex; many proposals have been made to simplify them. This is largely due to the fact that if only one use

were made of the bibliographic record, then the content designators could be reduced to those necessary for that one kind of product. For example, tags and fixed-field information are needed for information retrieval; tags and indicators for filing; subfield codes for printing abbreviated records; and so on. In 1972 a Council on Library Resources-sponsored committee, CEMBI (Conference to Explore Machine-readable Bibliographic Interchange), surveyed current users of the format to determine which coded information and which content designators might be dropped. The results showed that there was no single tag, indicator, or subfield code not required by some user. It appears that while a simplified format can be established for one use or user, a format capable of being used cooperatively for many kinds of uses must carry within it the information necessary for the manipulation.

At the same time, it is undeniable that to carry out the editing necessary to put a record into the full MARC format is an expensive undertaking beyond the capabilities of some small libraries. This poses a potential difficulty in the concept of developing an automated national union catalog. Obviously, if the catalog is to be in machine-readable form, contribution of holdings in machine-readable form will be economically desirable. The National Library of Canada (NLC) has addressed this problem through its Mini-MARC format.[21] The principle behind Mini-MARC is that the creating library is allowed (within limits) to substitute a fill character for a content designator. When a record is submitted to NLC, the headings will be compared against their authority file and if a match is found, the fully content-designated field will be substituted. Other fields may be upgraded at NLC. The Library of Congress is also working on an NUC reporting format that will be less complex.[22] This format would reduce the content designators to a basic set so that subsequent processing through a format recognition program would produce a reasonably complete record. This concept is still under review at the Library of Congress.

While the main criticism of the MARC formats within the United States is that of too much complexity, the criticism in Europe is that the U.S. MARC format is too simple to cover the needs of bibliographic processing adequately. Development of MARC formats outside the United States began with U.K. MARC developed by the British National Bibliography (BNB). A representative from the BNB spent several weeks in the United States in 1967 working with staff of the Library of Congress while the MARC II format was being devel-

oped. The resulting U.K. MARC format was very similar to U.S. MARC, differing largely in the assignment of more complex subfield codes in the heading and title fields.[23] Since that time, MARC formats have been developed in most of the Western European countries, Canada, Latin America, Australia, and Japan. The Canadian, Latin American and Spanish formats[24] are very similar to the U.S. format; the Australian and Danish formats[25] are based more on the U.K. format; the INTERMARC format,[26] which was developed by representatives from France, Belgium, the Netherlands, Switzerland, and Great Britain, is somewhat more complex than the U.K. format; the German format[27] bears little similarity to the other formats. Some idea of the differences in complexity may be seen in the fact that in the U.S. MARC format, the title field is broken into three subfields, while in the INTERMARC format it is subdivided into twelve.

Exchange of data in machine-readable form on the international scale is obviously very difficult, given the multiplicity of formats. A given user, such as the Library of Congress, would require a separate program to process the bibliographic information coming from each national agency. To deal with this situation, the IFLA Working Group on Content Designators was formed in 1973 under the joint auspices of the Committee on Cataloging and the Committee on Mechanization. The purpose of the committee was to establish an international format to be used in the international exchange of data between national agencies. It was assumed that each country would probably continue to use its own national format but would translate records into or from the international format for exchange purposes. The new format, called UNIMARC (Universal MARC), was developed during a series of meetings culminating in an open meeting held in Paris in October 1975 to review the first preliminary draft. A second preliminary draft was at a meeting in Brussels in April 1976, and a published provisional version will then be made available.

The basis of the UNIMARC format is the International Standard Bibliographic Description (ISBD). Full content designation is provided for all data elements called for in ISBD. The general principle has been established that the format itself will not require the presence of any fields (except control number and title) but that if a field is present, it must be fully content-designated. Coded information, in general, will be optional.

The MARC formats were designed primarily to serve the needs of the library community. At the same time, work has also been carried on to develop formats to handle bibliographic records in the scientific

and technical community. Because this community has different needs, it has long used different cataloging rules from those used by libraries, and this has been inevitably reflected in its formats. An example of such a format is that given as Appendix A.3 to the ANSI Z39.2 1971 format.[28] The most prominent format developed today by the scientific and technical community is the UNISIST *Reference Manual.*[29] A working group sponsored by the Council on Library Resources is currently developing a format for journal articles and technical reports. Early evidence indicates that this group is basing its work largely on the *Reference Manual,* with some attempt being made to make it at least partially compatible with the MARC formats. Lack of agreement between the library formats and the scientific and technical formats promises to make communication between these two communities difficult.

CODES

Most formats carry information both in natural-language form and in coded form. The natural-language form is used in the printing of bibliographic entries and includes such fields as title, edition, imprint, and headings. The coded information is used partly to save characters but primarily for retrieval purposes. The MARC formats contain a considerable amount of coded data to show such information as language of text, country of publication, and whether a book is a juvenile book, biography, festschrift, government publication, etc. Many of the code sets used are very short. For example, in the biography field, *a* = autobiography, *b* = biography, *c* = collective biography, *d* = contains biographical information, and "blank" = not biographical. Other code sets are more complex and have themselves become standardized.

The U.S. MARC formats use a three-character alphabetic language code developed by a committee made up of staff from the three national libraries[30] and maintained at the Library of Congress. Due to the lack of a well-established international code, the LC code is widely used by other countries in their MARC formats. It is likely, however, that ISO will begin work on a new version of such a code in the near future.

The same national library committee also established a country code to cover country of publication, and (in the case of the United States, Canada, and the United Kingdom) state or province of publication.[31] This code was based in part on the state code developed for post office use. In 1974, ISO issued a new country code[32] which was

adopted in 1976 as a U.S. standard by ANSI.[33] This code is compatible with the LC code in that one code can be translated into the other, but the code values differ. It is probable that the U.S. MARC formats will continue in the near future to use the LC country code, but the ISO code will be used in the UNIMARC format.

The Library of Congress has also developed a seven-character alphabetic code called the Geographic Area Code (GAC)[34] which is used to describe the subject content of the book. The GAC describes not only the country, but also the continent, region, and in some cases, the state. The GAC is used in the U.S., Canadian, and British systems.

Another code developed at the Library of Congress is the Chronological Coverage Code (CCC)[35] which is made up of two-character alphanumeric codes. B.C. dates are coded by century and A.D. dates by decade. Date ranges are shown by the juxtaposition of two two-character codes. This code is currently used by the National Library of Canada but has not yet been implemented at LC.

ITEM IDENTIFICATION

In the development of manual systems for information interchange, the necessity of a unique item identification number was recognized early. The LC card number was one of the first of these and is still one of the most widely used numbers today. Similar numbers are found in the various national bibliographies. In the 1960s, the British National Bibliography pioneered in the development of an International Standard Book Number (ISBN) to be assigned by publishers when a book was issued. The ISBN is a ten-character number in which the first part of the number describes the language area, the second part is a number assigned to the publisher or distributor, and the third part is an item number. It was hoped that the ISBN would solve the problem of an international numbering system for bibliographic use and, in fact, many of the major European countries and the United States now have ISBN systems and these numbers are carried in bibliographic records. Unfortunately, the ISBN as applied by publishers represents a stock item and there is often no one-to-one correspondence between these stock items and the bibliographic record. For example, there are eight ISBNs assigned to the 1970 edition of *Webster's Seventh New Collegiate Dictionary*: one ISBN is assigned to the blue pigskin binding, one to the black leather binding, etc. Moreover, publishers and distributors sometimes each assign different numbers to the same item.

The International Standard Serial Number (ISSN) was developed by ISO to serve a similar function for serials. Unlike the ISBN, each ISSN number is assigned by a national agency and relates to only one bibliographic record. This number shows more promise of being useful in the exchange of bibliographic information, but the system has been beset with problems of designating when one serial stops and a new one begins and of persuading publishers to participate.

METHODS OF BIBLIOGRAPHIC EXCHANGE

The original method of distribution of bibliographic information was via the mails on magnetic tape. This method requires conformity in format, character set, codes, and tape labels. Tape labels are data describing the records on the tape, giving information on the issuing agency, date of issuance, etc.; they are carried in machine-readable form at the beginning of the tape. The tape labels used in the MARC tapes conform to the ANSI standard for labels.[36]

Communication of bibliographic data via telephone or satellite has now become a reality. Several networks have been established for the central creation of a data base with on-line searching and updating capabilities possible from remote terminals. The largest of these systems is OCLC, which contains a data base of more than 2 million records and is used by more than 650 libraries. The core of the OCLC data base comes from the MARC Distribution Service, but member libraries are allowed to input records when the desired record is not already on the data base. The OCLC system provides catalog cards to member libraries and, if requested, tape files containing the member library's records. Problems in the OCLC system include the absence of an authority file, the lack of control of duplication of records, and in some cases, a less-than-full level of cataloging. OCLC has demonstrated a remarkable success in providing custom-produced cards for member libraries. It is also widely used for searching and interlibrary loan activities.

Other networks include the BALLOTS system developed at Stanford University (which now provides on-line services to other institutions) and the Washington State Library Network. The latter system includes an on-line authority file against which headings in new records are automatically verified.

All of these systems are designed to send information from the central computer to the user's display terminal. If desired, the display can be printed out on a printer. The Library of Congress and the

Research Library Group (RLG), consisting of the Yale, Harvard, Columbia, and New York Public (NYPL) libraries, have agreed to provide computer-to-computer service between the LC and NYPL computer systems. RLG members (initially NYPL and Columbia) will be able to search the LC MARC data base on-line via the NYPL computer, and selected records will then be sent over the wires to the NYPL computer to be used in the RLG system. This arrangement may serve as a pilot effort for future development of a national network in which user libraries search regional data bases and finally a national data base to obtain machine-readable records.

Many problems remain to be solved before such networks can become a reality. These include standardization of protocols for computer-to-computer communication, compatibility of indexes, and problems of updating the files. The question of protocol is essentially a technical problem and is now being addressed by the ALA/ISAD Telecommunications Committee. The problem of compatibility of indexes derives from the different index keys used in various systems. For example, OCLC uses an author-title key made up of three letters from the author's name and three letters from the title. BALLOTS uses key words in the author and title fields. In the ideal network, a user should be able to search the OCLC system and if no hit is found, the system should be able to transfer the search to another regional network. However, in this case, the OCLC six-character author-title key could not be automatically coverted to a BALLOTS keyword search key. The system of the future will probably require a common set of data elements which will be translated into the query terms used by a given system. Work in this area is currently being conducted at the Massachusetts Institute of Technology.

The problem of updating records may be even more difficult. If LC modifies a record and that record is held in four regional networks, will the LC update be sent to modify these records? If a user of the BALLOTS system modifies a record, will that modification be sent to all other networks holding that record? What will happen if the modifications made by a BALLOTS user and those made by an OCLC user conflict? These problems will have to be solved before a truly interactive network can be established.

EXCLUSIVE USE OF BIBLIOGRAPHIC RECORDS

MARC records distributed by the Library of Congress are in the public domain and are made available at cost plus 10 percent to users.

(The 10 percent is required by law.) The expense of developing automated systems and the relative ease with which automated data can be used to produce a multiplicity of products, however, have raised the question of exclusive right of use. The British Library has negotiated an agreement with the National Library of Canada under which each library will agree to limit the distribution of the other country's records to nonprofit educational and cultural agencies in their respective countries. The Library of Congress, on the other hand, has taken the position that it cannot and should not try to control distribution in the United States, and has therefore negotiated agreements with Canada, France and Australia under which records for each country's imprints will be distributed on a nonexclusive basis. This philosophy is also reflected in the recent agreement between the Council on Library Resources and OCLC under which serials in the CONSER (Conversion of Serials) project input, updated, or claimed by the participating CONSER libraries will be made available to CLR for distribution through the LC MARC Distribution Service.

In summary, it is clear that great strides have been taken in the development of standards necessary for the exchange of bibliographic data. Much work has been done, but new agreements requiring compromises by all involved will be necessary before the full potential for automated library networks can be realized.

References

1. Wigington, Ronald L., and Wood, James L. "Standardization Requirements of a National Program for Information Transfer," *Library Trends* 18:432-47, April 1970.

2. Schmierer, Helen F. "Bibliographic Standards." *In* Carlos A. Cuadra, ed. *Annual Review of Information Science and Technology*. Vol 10. Washington, D.C., American Society for Information Science, 1975, pp. 105-38.

3. "Standards for Library Automation and ISAD's Committee on Technical Standards for Library Automation (TESLA)," *Journal of Library Automation* 7:126-38, June 1974.

4. International Federation of Library Associations. *ISBD(M): International Standard Bibliographic Description for Single Volume and Multivolume Monographic Publications, 1974.* 1st standard ed. London, IFLA Committee on Cataloguing, 1974.

5. *Anglo-American Cataloging Rules. Chapter 6, Separately Published Monographs . . . (North American Text).* Chicago, ALA, 1974.

6. U.S. Library of Congress. Information Systems Office. *Format Recognition Process for MARC Records: A Logical Design.* Chicago, ALA Information Science and Automation Division, 1970.

7. *National Union Catalog: A Cumulative Author List Representing Library of Congress Printed Cards and Titles Reported by Other American Libraries.* Washington, D.C., Library of Congress, 1956- .

8. *Library of Congress Name Headings with References.* Washington, D.C., Library of Congress, 1974- .

9. U.S. Standards Institute. *USA Standard Code for Information Interchange.* USAS X3.4-1968. New York, U.S. Standards Institute, 1968.

10. U.S. Library of Congress. MARC Development Office. "Expanded Library Character Set." Revised Oct. 21, 1970. [Washington, D.C., MARC Development Office, Library of Congress, 1970.] (mimeographed)

11. International Organization for Standardization. "Extended Latin Alphabet Character Set for Bibliographic Use" (Draft International Standard ISO/DP 5426). TC46 Documentation, 1975. (Processed, unpaged); ———. "Greek Alphabet Character Set for Bibliographic Use" (Draft International Standard ISO/DIS 5428). TC46 Documentation, 1975. (Processed, unpaged); and ———. "Extended Cyrillic Alphabet Character Set for Bibliographic Use" (Draft International Standard ISO/DP 5427). TC46 Documentation, 1975. (Processed, unpaged.)

12. Buckland, Lawrence F. *The Recording of Library of Congress Bibliographical Data in Machine Form.* Washington, D.C., Council on Library Resources, 1965.

13. Avram, Henriette D., *et al. A Proposed Format for a Standardized Machine-readable Catalog Record; A Preliminary Draft* (U.S. Library of Congress Information Systems Office Planning Memorandum No. 3). Washington, D.C., Library of Congress, 1971.

14. Avram, Henriette D. *The MARC Pilot Project; Final Report on a Project Sponsored by the Council on Library Resources, Inc.* Washington, D.C., Library of Congress, 1968.

15. ———, *et al. The MARC II Format; A Communications Format for Bibliographic Data.* Washington, D.C., Information Systems Office, Library of Congress, 1968. Replaced and updated *by* U.S. Library of Congress. MARC Development Office. *Books: A MARC Format; Specifications for Magnetic Tapes Containing Monographic Catalog Records in the MARC II Format.* 5th ed. Washington, D.C., Library of Congress, 1972.

16. American National Standards Institute. *American National Standard for Bibliographic Information Interchange on Magnetic Tape* (Z39.2-1971). New York, ANSI, 1971; and International Organization for Standardization. *Documentation—Format for Bibliographic Information Interchange on Magnetic Tape* (ISO 2709). Geneva, 1973.

17. U.S. Library of Congress. MARC Development Office. *Serials: A MARC Format; Specifications for Magnetic Tapes Containing Catalog Records for Serials.* 2d ed. Washington, D.C., Library of Congress, 1974; ———. *Films: A MARC Format; Specifications for Magnetic Tapes Containing Catalog Records for Motion Pictures, Filmstrips, and Other Pictorial Media Intended for Projection.* Washington, D.C., Library of Congress, 1970; ———. *Maps: A MARC Format; Specifications for Magnetic Tapes Containing Catalog Records for Maps.* 2d ed. Washington, D.C., Library of Congress, 1976 (In press); ———. *Manuscripts: A MARC Format; Specifications for Magnetic Tapes Containing Catalog Records for Single Manuscripts or Manuscript Collections.* Washington,

D.C., Library of Congress, 1973; and _____. *Music: A MARC Format; Specifications for Magnetic Tapes Containing Catalog Records for Music Scores and Musical and Nonmusical Sound Recordings.* Washington, D.C., Library of Congress, 1976. (In press.)

18. _____. Composite MARC Format; A Tabular Listing of Content Designators Used in the MARC Formats. Washington, D.C., Library of Congress. (In process.)

19. _____. "Authorities: A MARC Format" (Draft-working document). Washington, D.C., Library of Congress, 1975.

20. _____. *In Process Materials: A MARC Format; Specifications for Magnetic Tapes Containing Records for Materials in the Process of Being Cataloged in the Library of Congress.* Washington, D.C., Library of Congress, 1976. (In press.)

21. National Library of Canada. Research and Planning Branch. Canadian MARC Office. *Canadian MARC Communication Format: Mini-MARC; or Canadian National Union Catalogue Accession Reporting Communication Format: Monographs and Serials.* Ottawa, National Library of Canada, 1975.

22. Weisbrod, David L. "An NUC Reporting Format in Machine-readable Form." Draft prepared for MARC Development Office. 1975.

23. British Library. Bibliographic Services Division. *U.K. MARC Manual.* 1st standard ed. London, British Library Bibliographic Services Division, 1975.

24. National Library of Canada. Research and Planning Branch. Canadian MARC Office. *Canadian MARC Communication Format: Monographs.* Ottawa, National Library of Canada, 1974; "Formato MARCAL: Libros." Bosquejo Preliminar. (Mid-Project Meeting on MARCAL Format, I MARCAL Working Paper No. 10.) Washington, D.C., 1976 (processed); and Spain. Comisaría Nacional de Bibliotecas. *IBERMARC: Formata para Monografías, Manual.* Madrid, 1976.

25. National Library of Australia. *Australian MARC Specification.* Canberra, National Library of Australia, 1973; and Thorlund, Gunner, *et al. Marcformatet i Praksis med 25 Eksempler* (Ebd-kontorets publikationer, 4). Kobenhavn, Forskningsbibliotekernes Edb-Kontor, 1974.

26. *INTERMARC (M), Format bibliographique d'échange pour les monographies: Manuel.* Paris, Bibliothéque nationale, 1975.

27. Deutsche Forschungsgemeinschaft. *Vorschlag für ein Maschinelles Austauschformat für Bibliotheken (MAB1).* München, 1973.

28. American National Standards Institute, *op. cit.*, pp. 15-32.

29. UNISIST/ICSU-AB Working Group on Bibliographic Descriptions. *Reference Manual for Machine-readable Bibliographic Descriptions.* M. D. Martin, comp. Paris, Unesco, 1974.

30. U.S. Library of Congress. MARC Development Office. "Revised List of Languages and Language Codes (Recommended by the Working Group on Bibliographical Codes)." Rev. Feb. 1973. Washington, D.C., Library of Congress, 1973. (processed)

31. _____. "Country of Publication Codes (Including Selected First Level Administrative Subdivisions)." Rev. Feb. 1973. Washington, D.C., Library of Congress, 1973. (processed)

32. International Organization for Standardization. *Codes for the Representation of Names of Countries* (ISO 3166). Geneva, ISO, 1974.

33. American National Standards Institute. *Structure for the Representation of Names of Countries of the World for Information Interchange* (Z39.27-1976). New York, ANSI, 1976.

34. U.S. Library of Congress. MARC Development Office. "Geographic Area Codes." Rev. Aug. 1973. Washington, D.C., Library of Congress, 1973. (processed)

35. "Chronological Coverage Code." *In* U.S. Library of Congress, MARC Development Office, *Books: A MARC Format, op. cit.*, pp. 95-97.

36. American National Standards Institute. *American National Standard Magnetic Tape Labels for Information Interchange* (X3.27-1969). New York, ANSI, 1969.

This Page Intentionally Left Blank